

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re U.S. Patent Application )

Applicant: Bita et al. )

Serial No. )

Filed: November 19, 2003 )

For: STORAGE CONTROL )  
APPARATUS AND CONTROL )  
METHOD THEREOF )

*I hereby certify that this paper is being deposited with the United States Postal Service as EXPRESS MAIL in an envelope addressed to: Mail Stop PATENT APPLICATION, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on this date.*

Nov. 19, 2003  
Date

  
Express Mail Label No.: EV032735480US

CLAIM FOR PRIORITY

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Dear Sir:

Applicants claim foreign priority benefits under 35 U.S.C. § 119 on the basis of the foreign application identified below:

Japanese Patent Application No. 2002-338901, filed November 22, 2002

A certified copy of the priority document is enclosed.

Respectfully submitted,

GREER, BURNS & CRAIN, LTD.

By



Patrick G. Burns

Registration No. 29,367

November 19, 2003  
300 South Wacker Drive  
Suite 2500  
Chicago, Illinois 60606  
Telephone: 312.360.0080  
Facsimile: 312.360.9315

3408.68745  
312.360.0080



日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 2 年 1 1 月 2 2 日  
Date of Application:

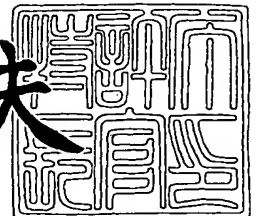
出 願 番 号                      特 願 2 0 0 2 - 3 3 8 9 0 1  
Application Number:  
[ST. 10/C]:                      [ J P 2 0 0 2 - 3 3 8 9 0 1 ]

出      願      人                      富 士 通 株 式 会 社  
Applicant(s):

2 0 0 3 年    8 月    1 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



出証番号    出証特 2 0 0 3 - 3 0 6 1 9 1 1

【書類名】 特許願

【整理番号】 0252773

【提出日】 平成14年11月22日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 13/12 340

【発明の名称】 ストレージ制御装置及びその制御方法

【請求項の数】 5

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

【氏名】 尾田 丈一

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

【氏名】 仲村 大也

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【代理人】

【識別番号】 100094514

【弁理士】

【氏名又は名称】 林 恒徳

【選任した代理人】

【識別番号】 100094525

【弁理士】

【氏名又は名称】 土井 健二

【手数料の表示】

【予納台帳番号】 030708

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9704944

【プルーフの要否】 要

**【書類名】 明細書****【発明の名称】 ストレージ制御装置及びその制御方法****【特許請求の範囲】**

**【請求項 1】** 依頼装置からのデータアクセス要求に応じて、ストレージ装置をアクセスするストレージ制御装置において、

第 1 のキャッシュメモリを有し、複数のストレージ装置の内、第 1 のストレージ装置を担当する第 1 のコントローラと、

第 2 のキャッシュメモリを有し、前記複数のストレージ装置の内、第 2 のストレージ装置を担当する第 2 のコントローラとを有し、

前記第 1 のコントローラは、前記第 2 のキャッシュメモリのミラー領域を管理する第 1 のミラー管理テーブルを有し、

前記第 2 のコントローラは、前記第 1 のキャッシュメモリのミラー領域を管理する第 2 のミラー管理テーブルを有し、

前記第 1 のコントローラが、前記依頼装置からデータ書込み要求を受けたことに応じて、前記第 1 のキャッシュメモリのリード／ライト領域のページを割り付けるとともに、前記第 1 のミラー管理テーブルを参照して、前記第 2 のキャッシュメモリのミラー領域の格納ページを獲得し、前記依頼装置からの書込みデータを、前記第 1 のキャッシュメモリのリード／ライト領域の割り付けられたページに書込み、且つ前記第 2 のキャッシュメモリのミラー領域の獲得ページにコピーする

ことを特徴とするストレージ制御装置。

**【請求項 2】** 依頼装置からのデータアクセス要求に応じて、一対のコントローラのいずれかが、ストレージ装置をアクセスするストレージ制御方法において、

前記一対の一方のコントローラが、前記依頼装置からデータ書込み要求を受けたことに応じて、前記一方のコントローラに設けられた第 1 のキャッシュメモリのリード／ライト領域のページを割り付けるステップと、

前記一方のコントローラに設けられ、前記他方のコントローラの第 2 のキャッシュメモリのミラー領域を管理する第 1 のミラー管理テーブルを参照して、前記

第2のキャッシュメモリのミラー領域の格納ページを獲得するステップと、

前記依頼装置からの書込みデータを、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書込むステップと、

前記書込み後に前記第2のキャッシュメモリのミラー領域の獲得ページにコピーするステップとを有する

ことを特徴とするストレージ制御方法。

【請求項3】 前記一方のコントローラは、前記第1及び第2のキャッシュメモリのサイズを相互に通知し、前記サイズに応じて、前記第1及び第2のキャッシュメモリのミラー領域を割り当て、前記第1及び第2のミラー管理テーブルを作成するステップを更に有する

ことを特徴とする請求項2のストレージ制御方法。

【請求項4】 前記一方のコントローラは、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書込まれたデータを、前記ストレージ装置にライトバックするステップと、

前記ライトバック完了により、前記第1のミラー管理テーブルの前記獲得ページを開放するステップを更に有する

ことを特徴とする請求項2のストレージ制御方法。

【請求項5】 前記一方のコントローラが、前記依頼装置から複数ページ分のデータ書込み要求を受けたことに応じて、前記コピー中に、前記第1のキャッシュメモリのリード／ライト領域の割り付けられた次ページに次ページのデータを書込むステップを更に有する

ことを特徴とする請求項2のストレージ制御方法。

#### 【発明の詳細な説明】

##### 【0001】

##### 【発明の属する技術分野】

本発明は、磁気ディスク等の物理ディスクを使用してデータを管理するストレージ制御装置及びその方法に関し、特に、コントローラを2重化したストレージ制御装置及びその方法に関する。

##### 【0002】

**【従来の技術】**

磁気ディスク、光磁気ディスク、光ディスク等の記憶媒体を利用したストレージ機器では、データ処理装置の要求で、記憶媒体を実アクセスする。データ処理装置が、大容量のデータを使用する場合には、複数のストレージ機器と制御装置とを備えたストレージ装置を利用する。

**【0003】**

このようなストレージ装置では、保存データの信頼性や、装置の信頼性を向上するため、冗長構成を採用している。図12は、従来技術の説明図である。

**【0004】**

図12に示すように、各々ホストに接続する一対のコントローラ100、110に、複数のストレージ機器（ディスク装置）150-1～150-nが接続されている。冗長構成のため、通常、ディスク装置150-1～150-nは、RAID（Redundant Array of Inexpensive (or Independent) Disks）というディスクの多重化構成を採用する。又、コントローラを2重化し、一対のコントローラ100、110で構成する。

**【0005】**

従って、通常は、両コントローラ100、110で、ディスク装置150-1～150-nを分担して制御するが、片方のコントローラ（例えば、100）が故障しても、他方のコントローラ（例えば、110）が、ディスク装置150-1～150-nを制御することができる。

**【0006】**

又、ホストアクセスのレスポンスを向上するため、コントローラ100、110内に、キャッシュメモリ120、130を設けている。このキャッシュメモリ120、130は、リード／ライト領域とミラー領域とを有し、リード／ライト領域には、自己が分担するディスク装置のデータの一部と、ホストからのライトデータを格納する。

**【0007】**

即ち、ホストからのリードアクセスに対し、対象データが、キャッシュメモリ120、130にステージングされていれば、実際に、ディスク装置をアクセス

することなく、キャッシュメモリのデータを読み出し、転送する。又、ホストからのライトアクセスに対しては、キャッシュメモリ 120、130 にライトデータを書き込んで、書き込み完了とする。キャッシュメモリ 120、130 のライトデータは、処理の空き時間に、対象ディスク装置にライトバックされる。

#### 【0008】

一方、ミラー領域は、前述のように、コントローラの 2 重化のため、一方のコントローラ 100 又は 110 が、自己のキャッシュメモリ 120 又は 130 に書き込んだライトデータを、他方のコントローラ 110 又は 100 のキャッシュメモリ 130 又は 120 にコピーしておくためのものである。これにより、前述のように、一方のコントローラが故障しても、他方のコントローラが、ミラー領域に、一方のコントローラのライトバック前のライトデータを保持しているため、データの消失を防止できる。

#### 【0009】

従来、このライトデータのミラーリングは、次のように行われていた。ここでは、コントローラ 100 にライトアクセスが到来した場合を説明するが、コントローラ 110 にライトアクセスが到来した場合も同様である。

#### 【0010】

コントローラ 100 が、ホストからデータ書き込み要求を受けると (①)、キャッシュメモリ 120 の書き込みページの割り付けを行う (②)。次に、コントローラ 100 は、コントローラ 110 に、キャッシュメモリ 130 のミラーページの獲得要求を送信する (③)。コントローラ 110 は、キャッシュメモリ 130 のミラーページの割り付けを行った後 (④)、コントローラ 100 にミラーページの獲得を応答する (⑤)。

#### 【0011】

次に、ホストからのデータを、キャッシュメモリ 120 の割り付けたページに書き込む (⑥)。キャッシュメモリ 120 の書き込まれたデータを、コントローラ 110 の割り付けられたミラーページにコピーする (⑦)。コピー完了後、コントローラ 100 は、ホストにデータ書き込み完了報告を通知する (⑧)。

#### 【0012】



このようにして、ライトデータのミラーリングを行っていた（例えば、特許文献1参照）。

**【0013】**

**【特許文献1】**

特開 2002-132453号公報

**【0014】**

**【発明が解決しようとする課題】**

従来技術では、ミラーページの獲得のため、コントローラ100、120が、獲得依頼、獲得応答のメッセージのやりとりが必要であり、コントローラ間のハードの負荷が大きい。このため、ライトデータのミラーリング時の性能向上が困難である。

**【0015】**

又、ミラーページ獲得のため、相手コントローラ（図12では、コントローラ110）のプログラム動作が必要となり、自コントローラの処理にもかかわらず、相手コントローラの負荷が生じ、2つのコントローラとしての性能向上が困難である。

**【0016】**

従って、本発明の目的は、ライトデータのミラーリングを、コントローラ間の負荷を軽減して実行するためのストレージ制御装置及び制御方法を提供するにある。

**【0017】**

又、本発明の他の目的は、相手コントローラの負荷を軽減して、ライトデータのミラーリングを実行するためのストレージ制御装置及びその制御方法を提供するにある。

**【0018】**

更に、本発明の他の目的は、ライトデータのミラーリングを高速に実行するためのストレージ制御装置及びその制御方法を提供するにある。

**【0019】**

**【課題を解決するための手段】**

この目的の達成のため、本発明のストレージ制御装置は、第1のキャッシュメモリを有し、複数のストレージ装置の内、第1のストレージ装置を担当する第1のコントローラと、第2のキャッシュメモリを有し、前記複数のストレージ装置の内、第2のストレージ装置を担当する第2のコントローラとを有し、前記第1のコントローラは、前記第2のキャッシュメモリのミラー領域を管理する第1のミラー管理テーブルを有し、前記第2のコントローラは、前記第1のキャッシュメモリのミラー領域を管理する第2のミラー管理テーブルを有し、前記第1のコントローラが、前記依頼装置からデータ書き込み要求を受けたことに応じて、前記第1のキャッシュメモリのリード／ライト領域のページを割り付けるとともに、前記第1のミラー管理テーブルを参照して、前記第2のキャッシュメモリのミラー領域の格納ページを獲得し、前記依頼装置からの書き込みデータを、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書き込み、且つ前記第2のキャッシュメモリのミラー領域の獲得ページにコピーする。

#### 【0020】

又、本発明のストレージ制御方法は、依頼装置からのデータアクセス要求に応じて、一対のコントローラのいずれかが、ストレージ装置をアクセスするストレージ制御方法において、前記一対の一方のコントローラが、前記依頼装置からデータ書き込み要求を受けたことに応じて、前記一方のコントローラに設けられた第1のキャッシュメモリのリード／ライト領域のページを割り付けるステップと、前記一方のコントローラに設けられ、前記他方のコントローラの第2のキャッシュメモリのミラー領域を管理する第1のミラー管理テーブルを参照して、前記第2のキャッシュメモリのミラー領域の格納ページを獲得するステップと、前記依頼装置からの書き込みデータを、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書き込むステップと、前記書き込み後に前記第2のキャッシュメモリのミラー領域の獲得ページにコピーするステップとを有する。

#### 【0021】

本発明では、相手コントローラのキャッシュメモリのミラー領域のミラー管理テーブルを自己のコントローラが保有し、ミラーページ獲得を行うことにより、コントローラ間でのミラーページ獲得メッセージのやり取りが不要となり、コン

トローラ間のハード、特に、CPUの負荷を軽減して、ライトデータのミラーリングが可能となる。しかも、相手コントローラのプログラム動作を不要とでき、コントローラ自体の性能を向上できる。

#### 【0022】

又、本発明のストレージ制御装置は、好ましくは、前記第1及び前記第2のコントローラは、前記第1及び第2のキャッシュメモリのサイズを相互に通知し、前記サイズに応じて、前記第1及び第2のキャッシュメモリのミラー領域を割り当て、前記第1及び第2のミラー管理テーブルを作成する。

#### 【0023】

又、本発明のストレージ制御方法は、好ましくは、前記一対のコントローラは、前記第1及び第2のキャッシュメモリのサイズを相互に通知し、前記サイズに応じて、前記第1及び第2のキャッシュメモリのミラー領域を割り当て、前記第1及び第2のミラー管理テーブルを作成するステップを更に有する。

#### 【0024】

この本発明の態様では、相手コントローラのキャッシュメモリのミラー領域を割り当て、ミラー管理テーブルを作成することにより、相手コントローラのキャッシュメモリのサイズに応じたミラー領域を設定でき、コントローラ間でのミラーページ獲得メッセージのやり取りが不要となる。

#### 【0025】

又、本発明のストレージ制御装置では、好ましくは、前記第1のコントローラは、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書込まれたデータを、前記ストレージ装置にライトバックした後、前記第1のミラー管理テーブルの前記獲得ページを開放する。

#### 【0026】

又、本発明のストレージ制御方法は、好ましくは、前記一方のコントローラは、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書込まれたデータを、前記ストレージ装置にライトバックするステップと、前記ライトバック完了により、前記第1のミラー管理テーブルの前記獲得ページを開放するステップを更に有する。

**【0027】**

この本発明の態様では、ライトバック時に、ミラー領域を管理する管理テーブルで、ミラーページを開放するので、他方のコントローラの手間がなく、他方のコントローラのミラー開放が可能となる。

**【0028】**

又、本発明のストレージ制御装置は、好ましくは、前記第1のコントローラがデグレードした時に、前記第2のコントローラが、前記第1のコントローラが担当するストレージ装置も担当し、且つ前記第2のキャッシュメモリのミラー領域のコピーページを前記リード／ライト領域にリンクする。

**【0029】**

又、本発明のストレージ制御方法は、好ましくは、前記一方のコントローラがデグレードした時に、前記他方のコントローラが、前記一方のコントローラが担当するストレージ装置も担当し、且つ前記第2のキャッシュメモリのミラー領域のコピーページを第2のキャッシュメモリのリード／ライト領域にリンクするステップとを更に有する。

**【0030】**

これにより、一方のコントローラがデグレードした時に、他方のコントローラは、ミラーデータを引き継ぎ、代行して処理することができる。

**【0031】**

又、本発明のストレージ制御装置及び方法は、好ましくは、前記第2のコントローラは、前記第2のキャッシュメモリのミラー領域へのリード／ライト処理を禁止する。これにより、デグレードしたコントローラが、立ち上がった時に、ミラー領域を確保できる。

**【0032】**

又、本発明のストレージ制御装置又は方法は、好ましくは、前記第1のコントローラが、前記依頼装置から複数ページ分のデータ書込み要求を受けたことに応じて、前記第1のキャッシュメモリのリード／ライト領域のページを複数割り付けるとともに、前記第1のミラー管理テーブルを参照して、前記第2のキャッシュメモリのミラー領域の複数の格納ページを獲得し、前記依頼装置からの書込み

データを、前記第 1 のキャッシュメモリのリード／ライト領域の割り付けられたページに書込み、且つ前記第 2 のキャッシュメモリのミラー領域の獲得ページにコピーし、前記コピー中に、前記第 1 のキャッシュメモリのリード／ライト領域の割り付けられた次ページに次ページのデータを書込む。

#### 【 0 0 3 3 】

これにより、連続ページの書込み処理を、ミラーリングしながら、高速化できる。

#### 【 0 0 3 4 】

又、本発明のストレージ制御装置又は方法は、好ましくは、前記第 1 及び第 2 のコントローラの各々は、前記キャッシュメモリ及び前記ストレージ装置を制御する制御ユニットと、前記コントローラ間での通信を行うためのノードチャネル回路とを有する。

#### 【 0 0 3 5 】

又、本発明のストレージ制御装置又は方法は、好ましくは、前記制御ユニットは、前記ノードチャネル回路に、前記第 1 のキャッシュメモリのリード／ライト領域の割り付けられたページのデータを、前記第 2 のキャッシュメモリのミラー領域の獲得ページに DMA 転送することを指示し、前記コピーを行う。

#### 【 0 0 3 6 】

このため、制御ユニットのコピー負荷が軽減でき、ミラーリング処理を、最小の負荷で実行できる。

#### 【 0 0 3 7 】

#### 【発明の実施の形態】

以下、本発明の実施の形態を、ストレージシステム、ライト処理、ライトバック処理、デグレード処理、他の実施の形態の順で説明する。

#### 【 0 0 3 8 】

#### 〔ストレージシステム〕

図 1 は、本発明の一実施の形態のストレージシステムの構成図であり、磁気ディスクを使用した R A I D (Redundant Arrays of Inexpensive Disk) システムを示す。図 1 に示すように、ストレージシステムは、一対の磁気ディスクコント

ローラ（以下、コントローラという） 1、2 と、この一対のコントローラ 1、2 にライン 1 1、1 2 で接続された多数の磁気ディスク装置 5 0 - 1 ~ 5 0 - m、5 2 - 1 ~ 5 2 - n とからなる。

#### 【 0 0 3 9 】

コントローラ 1、2 は、直接又はネットワーク機器を介し、ホストやサーバーに接続され、ホストやサーバーの大量のデータを、R A I D ディスクドライブ（磁気ディスク装置）へ高速かつ、ランダムに読み書きが出来るシステムである。一対のコントローラ 1、2 は、同一の構成を有し、C A（Channel Adapter）1 1、1 2、2 1、2 2 と、C M（Centralized Module）1 0、1 5 ~ 1 9、2 0、2 5 ~ 2 9 と、D A（Device Adapter）1 3、1 4、2 3、2 4 のファンクションモジュールによって構成されている。

#### 【 0 0 4 0 】

C A（Channel Adapter）1 1、1 2、2 1、2 2 は、ホストを結ぶホスト・インタフェースの制御をつかさどる回路であり、例えば、ファイバーチャネル回路（F C）と D M A（Direct Memory Access）回路等で構成される。D A（Device Adapter）1 3、1 4、2 3、2 4 は、ディスクデバイス 5 0 - 1 ~ 5 0 - m、5 2 - 1 ~ 5 2 - m を制御するため、ディスクデバイスとコマンド、データのやり取りを行う回路であり、例えば、ファイバーチャネル回路（F C）と D M A 回路等で構成される。

#### 【 0 0 4 1 】

C M（Centralized Module）は、C P U 1 0、2 0 と、ブリッジ回路 1 7、2 7 と、メモリー（R A M）1 5、2 5 と、フラッシュメモリ 1 9、2 9 と、I O ブリッジ回路 1 8、2 8 とを有する。メモリー 1 5、2 5 は、バッテリーでバックアップされ、その一部が、キャッシュメモリ 1 6、2 6 として使用される。

C P U 1 0、2 0 は、ブリッジ回路 1 7 を介し、メモリー 1 5、2 5、フラッシュメモリ 1 9、2 9、I O ブリッジ回路 1 8、2 8 に接続される。このメモリー 1 5、2 5 は、C P U 1 0、2 0 のワーク領域に使用され、フラッシュメモリ 1 9、2 9 は、C P U 1 0、2 0 が実行するプログラムを格納する。このプログラムとして、B I O S（Basic Input/Output System）、ファイルアクセスプロ

グラム（リード／ライトプログラム）、RAID管理プログラム等を格納する。CPU10, 20は、このプログラムを実行し、リード／ライト処理、RAID管理処理等を実行する。

#### 【0042】

PCI (Personal Computer Interface) バス31は、CA11, 12, 21, 22と、DA13, 14, 23, 24とを接続するとともに、IOブリッジ回路18を介し、CPU10, 20、メモリー15, 25を接続する。更に、PCIバス31には、PCI－ノードリンクブリッジ回路30, 40が接続される。コントローラ1のPCI－ノードリンクブリッジ回路30は、コントローラ2のPCI－ノードリンクブリッジ回路40と接続され、コントローラ1, 2間のコマンド、データの交信を行う。

#### 【0043】

コントローラ1は、例えば、ディスク装置50-1～50-mを担当し、コントローラ2は、例えば、ディスク装置52-1～52-nを担当する。図1では、ディスク装置50-1～50-mと、52-1～52-nとが、RAID5の構成を有する。同様に、RAID1, RAID0+1の構成も可能である。

#### 【0044】

キャッシュメモリ16, 26は、各々、担当するディスク装置のデータの一部を格納し、ホストからのライトデータを格納する。CPU10, 20は、CA11, 12, 21, 22を介しホストからのリード要求を受けて、キャッシュメモリ16, 26を参照し、物理ディスクへのアクセスが必要かを判定し、必要であれば、ディスクアクセス要求をDA13, 14, 23, 24に要求する。又、CPU10, 20は、ホストからのライト要求を受けて、ライトデータをキャッシュメモリ16, 26に書込み、且つ内部でスケジュールされるライトバック等をDA13, 14, 23, 24に要求する。

#### 【0045】

##### [ライト処理]

次に、キャッシュメモリのミラー領域割り付け処理及びライト処理を、図2乃至図6で説明する。先ず、キャッシュメモリのミラー領域割り付け処理を、図2

乃至図4で説明する。図2は、本発明の一実施の形態のキャッシュメモリのミラー領域割り付け処理フロー図、図3は、図2のミラー領域割り付けの説明図、図4は、図3のミラー管理テーブルの説明図である。

#### 【0046】

図3に示すように、コントローラ1のキャッシュメモリ16は、コントローラ1のリード／ライト領域16-1と、コントローラ2のミラー領域16-2と、ミラー管理テーブル160とからなる。又、コントローラ2のキャッシュメモリ26は、コントローラ2のリード／ライト領域26-1と、コントローラ1のミラー領域26-2と、ミラー管理テーブル260とからなる。

#### 【0047】

ミラー管理テーブル160は、キャッシュメモリ26のミラー領域26-2の管理のための情報を記憶し、ミラー管理テーブル260は、キャッシュメモリ16のミラー領域16-2の管理のための情報を記憶する。図4に示すように、ミラー管理テーブル160、260は、ミラー領域として割り当てられた各ページ1～xと、各ページの次の空きページへのリンク情報とからなるページ管理テーブルMMTと、未使用のリンク先頭空きページを示す未使用ミラーポインタMPとを有する。

#### 【0048】

図3を参照して、図2により、ミラー領域割り付け処理を説明する。

#### 【0049】

(S10) 装置の立上げ時に、ミラー領域割り付け処理を実行する。

#### 【0050】

(S12) 先ず、コントローラ1又は2のCPU10又は20は、自己のキャッシュメモリ16又は26のサイズ（ページ数）を、PCI-ノードブリッジ回路30、40を介し、相手のコントローラ2又は1のCPU20又は10に通知する。

#### 【0051】

(S14) 通知されたコントローラ2又は1のCPU20又は10は、相手のコントローラ1又は2のキャッシュメモリ16又は26のキャッシュサイズの3



割分を、ミラー領域 16-1 又は 26-1 を割り当て、このミラーサイズのミラー管理テーブル 160 又は 260 を作成する。

#### 【0052】

例えば、キャッシュメモリ 16 の容量が、100 ページとし、キャッシュメモリ 26 の容量が、110 ページとすると、キャッシュメモリ 16 の 33 ページ分が、ミラー領域 16-2 に割り当てられ、キャッシュメモリ 26 の 30 ページ分が、ミラー領域 26-2 に割り当てられる。

#### 【0053】

従って、CPU 10 は、33 ページのミラー管理テーブル 160 を、CPU 20 は、30 ページのミラー管理テーブル 260 を作成する。この 3 割分は、例示であり、2 割等他の値を使用できる。

#### 【0054】

このように、ミラー領域を固定にすることにより、コントローラ間の獲得要求を無くすことができる。但し、担当する物理ディスクの容量等により、各コントローラのキャッシュサイズが異なる場合があるため、相互にキャッシュサイズを通知し、その固定割合をミラー領域に割り当てる。このため、キャッシュサイズの相違にも対応できる。又、キャッシュサイズが同一であれば、キャッシュサイズの相互通知は、不要である。

#### 【0055】

次に、図 5 を参照して、図 6 によりライト処理を説明する。尚、図 5 では、コントローラ 1 が、ホストからデータ書込み要求を受けた場合を説明するが、コントローラ 2 が、ホストからデータ書込み要求を受けた場合も同様である。

#### 【0056】

(S20) コントローラ 1 の CPU 10 は、CA11, 12 を介しホストからデータ書込み要求を受ける (①)。

#### 【0057】

(S22) CPU 10 は、キャッシュメモリ 16 のリード/ライト領域 16-1 の書込みページを割り付ける (②)。

#### 【0058】

(S24) CPU10は、ミラー管理テーブル160のミラーページ獲得要求を発する(③)。

**【0059】**

(S26) CPU10は、ミラー管理テーブル160のミラーページを獲得する。即ち、図4の未使用ミラーポインタMPに示す先頭空きページをミラーページとして、CPU10に通知する。そして、未使用ミラーポインタMPを、その通知したミラーページの次ページに更新し、ページ管理テーブルMMTのリンク情報を更新する(④)。

**【0060】**

(S28) データ書込み可能状態となり、コントローラ1のCPU10は、CA11, 12をするホストから書込みデータを、ステップS22で割り付けた書込みページに書込む(⑤)。即ち、CPU10は、CA11, 12に、リード/ライト領域16-1の書込みページへの書込みデータのDMA転送を指示する。

**【0061】**

(S30) CPU10は、書込みページに書込みデータを書き込んだ後、書込みページのデータを、コントローラ2のキャッシュメモリ26のミラー領域26-2のミラーページへコピーする(⑥)。即ち、CPU10は、PCI-ノードブリッジ回路30へ、書込みページのデータを、コントローラ2のキャッシュメモリ26のミラー領域26-2のミラーページへDMA転送を指示する。これにより、キャッシュメモリ16の書込みページのデータが、PCI-ノードブリッジ回路30, 40を介し、キャッシュメモリ26のミラーページにコピーされる。

**【0062】**

(S32) コピー完了後、CPU10は、CA11, 12を介しホストに、データ書込完了報告を通知する(⑦)。

**【0063】**

このようにして、相手コントローラのキャッシュメモリのミラー領域を割り当て、ミラー管理テーブルを作成することにより、コントローラ間でのミラーページ獲得メッセージのやり取りが不要となり、コントローラ間のハード、特に、C

P U の負荷を軽減して、ライトデータのミラーリングが可能となる。しかも、相手コントローラのプログラム動作を不要とでき、コントローラ自体の性能を向上できる。

#### 【0064】

##### [ライトバック処理]

次に、キャッシュメモリに書込まれたライトデータの物理ディスクへの書込みであるライトバック処理を説明する。図7は、本発明の一実施の形態のライトバック処理フロー図、図8は、ライトバック処理の説明図である。図8を参照して、図7により、ライトバック処理を説明する。尚、図8では、コントローラ1のキャッシュメモリ16のライトバックを説明するが、コントローラ2のキャッシュメモリ26のライトバックも同様である。

#### 【0065】

(S40) ライトバックは、ホストアクセス処理のバックグラウンドで所定のスケジュールで実行する。CPU10は、キャッシュメモリ16のリード／ライト領域16-1の対象ページ（ローカルページという）のライトバックを、DA13, 14に指示する。DA13, 14は、これにより、ローカルページのデータを読み出し、対応する物理ディスク50-1～50-mに、書込む(①)。

#### 【0066】

(S42) 書込み完了後、DA13, 14は、ライトバック完了を、CPU10に通知する。

#### 【0067】

(S44) CPU10は、このローカルページに対応するキャッシュメモリ26のミラーページのCBE (Cache Block Element: キャッシュブロック構成) に、初期化データ（例えば、a11「0」）をコピーする(②)。これにより、ミラーページの属性はフリーとなる。尚、ミラーページでは、そのCBEは、ミラーページ構成であることが書き込まれている。

#### 【0068】

(S46) この初期化コピー後、CPU10は、ミラー管理テーブル160の該当ミラーページの開放要求を発し(③)、ミラーページを開放する(④)。即

ち、図4の未使用ミラーポインタMPに示す先頭空きページをその開放するミラーページに更新し、ページ管理テーブルMMTのその開放ミラーページのリンク情報を、今まで先頭であった空きページに更新する。

#### 【0069】

(S48) その後、当該ローカルページをミラー状態から切り離す。即ち、ローカルページのCBEのミラー情報を初期化する。これにより、ライトバック処理を終了する。

#### 【0070】

このように、ライトバックにより、ミラー管理テーブル160からミラーページを開放する。又、ミラーページのCBEを操作するため、簡単に相手コントローラのキャッシュメモリのミラーページの状態を開放状態に変更できる。

#### 【0071】

##### [デグレート処理]

次に、片方のコントローラがデグレート（障害発生）した時の処理を、図9により説明する。ここでは、コントローラ1がデグレートした例で説明するが、コントローラ2がデグレートした場合も同様である。

#### 【0072】

コントローラ1がデグレートした場合に、コントローラ2は、自己の担当する物理ディスク52-1～52-nの他に、コントローラ1が担当していた物理ディスク50-1～50-mも担当する。従って、コントローラ2は、キャッシュメモリ26のミラー領域26-2（コントローラ1のミラーデータ）を、全て引き継がなければならない。

#### 【0073】

このため、図9に示すように、コントローラ2のCPU20は、キャッシュメモリ26のミラー領域26-2の全てのページのCBEをチェックし、ミラーページ属性のミラーページを、ハッシュテーブルにより、ローカルリンク（リード／ライト領域26-1のリンク）に組み込む。

#### 【0074】

これにより、コントローラ2は、コントローラ1が担当していた物理ディスク

50-1～50-mへのライトバックを含むリード／ライトアクセスを実行できる。

【0075】

又、ミラー領域26-2は、新規のリード／ライトの使用を禁止する。これにより、デグレードしたコントローラ1が、立ち上がった時に、コントローラ1のミラー領域26-2を確保できる。

【0076】

[他の実施の形態]

次に、本発明のライト処理の他の実施の形態を、図11を参照して、図10により説明する。尚、図11では、コントローラ1が、ホストからデータ書込み要求を受けた場合を説明するが、コントローラ2が、ホストからデータ書込み要求を受けた場合も同様である。

【0077】

(S50) コントローラ1のCPU10は、CA11, 12を介しホストから複数ページのデータ書込み要求を受ける(①)。

【0078】

(S52) CPU10は、キャッシュメモリ16のリード／ライト領域16-1に、複数の書込みページを割り付ける(②)。

【0079】

(S54) CPU10は、ミラー管理テーブル160の複数のミラーページ獲得要求を発する(③)。

【0080】

(S56) CPU10は、ミラー管理テーブル160の複数のミラーページを獲得する。即ち、図4の未使用ミラーポインタMPに示す先頭空きページをミラーページの先頭として、CPU10に通知する。そして、未使用ミラーポインタMPを、その通知したミラーページから複数ページ分、次ページに更新し、ページ管理テーブルMMTのリンク情報を更新する(④)。

【0081】

(S58) データ書込み可能状態となり、コントローラ1のCPU10は、C

A11, 12をするホストから1ページ目の書込みデータを、ステップS52で割り付けた先頭書込みページに書込む(⑤)。即ち、CPU10は、CA11, 12に、リード/ライト領域16-1の先頭書込みページへの書込みデータのDMA転送を指示する。

#### 【0082】

(S60) CPU10は、書込みページに書込みデータのDMA転送を指示し、書込み終了後、書込みページのデータを、コントローラ2のキャッシュメモリ26のミラー領域26-2のミラーページへコピー開始する(⑥)。即ち、CPU10は、PCI-ノードブリッジ回路30へ、書込みページのデータを、コントローラ2のキャッシュメモリ26のミラー領域26-2のミラーページへDMA転送を指示する。これにより、キャッシュメモリ16の書込みページのデータが、PCI-ノードブリッジ回路30, 40を介し、キャッシュメモリ26のミラーページにコピーされる。

#### 【0083】

(S62) CPU10は、コピー開始ページが最終ページかを判定する。

#### 【0084】

(S64) コピー開始ページが最終ページであるときは、コピー完了後、CPU10は、CA11, 12を介しホストに、データ書込完了報告を通知する(⑦)。

#### 【0085】

(S66) 一方、コピー開始ページが最終ページでない時は、コントローラ1のCPU10は、CA11, 12をするホストから次ページ目の書込みデータを、ステップS52で割り付けた次書込みページに書込む。即ち、CPU10は、CA11, 12に、リード/ライト領域16-1の次書込みページへの書込みデータのDMA転送を指示する。書込みページへの書込みと、コピーページのコピーが終了すると、ステップS60に戻る。

#### 【0086】

このようにして、相手コントローラのキャッシュメモリのミラー領域を割り当て、ミラー管理テーブルを作成することにより、コントローラ間でのミラーペー

ジ獲得メッセージのやり取りが不要となり、コントローラ間のハード、特に、CPUの負荷を軽減して、ライトデータのミラーリングが可能となる。しかも、相手コントローラのプログラム動作を不要とでき、コントローラ自体の性能を向上できる。

#### 【0087】

更に、複数のページにまたがるデータ書込み動作の場合に、データ書込み動作とデータコピー動作が、ハードウェアのデータ転送動作を伴い、ファーム動作に比較して遅いから、データコピー動作中に、次のキャッシュメモリのローカルページへのデータ書込みを動作させる。これにより、ミラーリングを伴うライト動作を高速化できる。

#### 【0088】

前述の実施の形態では、図1のような冗長構成のRAIDで説明したが、これ以外の冗長構成のストレージシステムに適用できる。又、物理ディスクは、磁気ディスク、光ディスク、光磁気ディスク、各種のストレージデバイスを適用できる。

#### 【0089】

以上、本発明を実施の形態により説明したが、本発明の趣旨の範囲内において、本発明は、種々の変形が可能であり、本発明の範囲からこれらを排除するものではない。

#### 【0090】

(付記1) 依頼装置からのデータアクセス要求に応じて、ストレージ装置をアクセスするストレージ制御装置において、第1のキャッシュメモリを有し、複数のストレージ装置の内、第1のストレージ装置を担当する第1のコントローラと、第2のキャッシュメモリを有し、前記複数のストレージ装置の内、第2のストレージ装置を担当する第2のコントローラとを有し、前記第1のコントローラは、前記第2のキャッシュメモリのミラー領域を管理する第1のミラー管理テーブルを有し、前記第2のコントローラは、前記第1のキャッシュメモリのミラー領域を管理する第2のミラー管理テーブルを有し、前記第1のコントローラが、前記依頼装置からデータ書込み要求を受けたことに応じて、前記第1のキャッシュ

メモリのリード／ライト領域のページを割り付けるとともに、前記第 1 のミラー管理テーブルを参照して、前記第 2 のキャッシュメモリのミラー領域の格納ページを獲得し、前記依頼装置からの書込みデータを、前記第 1 のキャッシュメモリのリード／ライト領域の割り付けられたページに書込み、且つ前記第 2 のキャッシュメモリのミラー領域の獲得ページにコピーすることを特徴とするストレージ制御装置。

#### 【 0 0 9 1 】

（付記 2）前記第 1 及び前記第 2 のコントローラは、前記第 1 及び第 2 のキャッシュメモリのサイズを相互に通知し、前記サイズに応じて、前記第 1 及び第 2 のキャッシュメモリのミラー領域を割り当て、前記第 1 及び第 2 のミラー管理テーブルを作成することを特徴とする付記 1 のストレージ制御装置。

#### 【 0 0 9 2 】

（付記 3）前記第 1 のコントローラは、前記第 1 のキャッシュメモリのリード／ライト領域の割り付けられたページに書込まれたデータを、前記ストレージ装置にライトバックした後、前記第 1 のミラー管理テーブルの前記獲得ページを開放することを特徴とする付記 1 のストレージ制御装置。

#### 【 0 0 9 3 】

（付記 4）前記第 1 のコントローラがデグレードした時に、前記第 2 のコントローラが、前記第 1 のコントローラが担当するストレージ装置も担当し、且つ前記第 2 のキャッシュメモリのミラー領域のコピーページを前記リード／ライト領域にリンクすることを特徴とする付記 1 のストレージ制御装置。

#### 【 0 0 9 4 】

（付記 5）前記第 2 のコントローラは、前記第 2 のキャッシュメモリのミラー領域へのリード／ライト処理を禁止することを特徴とする付記 4 のストレージ制御装置。

#### 【 0 0 9 5 】

（付記 6）前記第 1 のコントローラが、前記依頼装置から複数ページ分のデータ書込み要求を受けたことに応じて、前記第 1 のキャッシュメモリのリード／ライト領域のページを複数割り付けるとともに、前記第 1 のミラー管理テーブルを



参照して、前記第2のキャッシュメモリのミラー領域の複数の格納ページを獲得し、前記依頼装置からの書込みデータを、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書込み、且つ前記第2のキャッシュメモリのミラー領域の獲得ページにコピーし、前記コピー中に、前記第1のキャッシュメモリのリード／ライト領域の割り付けられた次ページに次ページのデータを書込むことを特徴とする付記1のストレージ制御装置。

#### 【0096】

(付記7) 前記第1及び第2のコントローラの各々は、前記キャッシュメモリ及び前記ストレージ装置を制御する制御ユニットと、前記コントローラ間での通信を行うためのノードチャネル回路とを有することを特徴とする付記1のストレージ制御装置。

#### 【0097】

(付記8) 前記制御ユニットは、前記ノードチャネル回路に、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページのデータを、前記第2のキャッシュメモリのミラー領域の獲得ページにDMA転送することを指示し、前記コピーを行うことを特徴とする付記7のストレージ制御装置。

#### 【0098】

(付記9) 依頼装置からのデータアクセス要求に応じて、一対のコントローラのいずれかが、ストレージ装置をアクセスするストレージ制御方法において、前記一対の一方のコントローラが、前記依頼装置からデータ書込み要求を受けたことに応じて、前記一方のコントローラに設けられた第1のキャッシュメモリのリード／ライト領域のページを割り付けるステップと、前記一方のコントローラに設けられ、前記他方のコントローラの第2のキャッシュメモリのミラー領域を管理する第1のミラー管理テーブルを参照して、前記第2のキャッシュメモリのミラー領域の格納ページを獲得するステップと、前記依頼装置からの書込みデータを、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書込むステップと、前記書込み後に前記第2のキャッシュメモリのミラー領域の獲得ページにコピーするステップとを有することを特徴とするストレージ制御方法。

**【0099】**

(付記10) 前記一対のコントローラは、前記第1及び第2のキャッシュメモリのサイズを相互に通知し、前記サイズに応じて、前記第1及び第2のキャッシュメモリのミラー領域を割り当て、前記第1及び第2のミラー管理テーブルを作成するステップを更に有することを特徴とする付記9のストレージ制御方法。

**【0100】**

(付記11) 前記一方のコントローラは、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページに書込まれたデータを、前記ストレージ装置にライトバックするステップと、前記ライトバック完了により、前記第1のミラー管理テーブルの前記獲得ページを開放するステップを更に有することを特徴とする付記9のストレージ制御方法。

**【0101】**

(付記12) 前記一方のコントローラがデグレードした時に、前記他方のコントローラが、前記一方のコントローラが担当するストレージ装置も担当し、且つ前記第2のキャッシュメモリのミラー領域のコピーページを第2のキャッシュメモリのリード／ライト領域にリンクするステップとを更に有することを特徴とする付記9のストレージ制御方法。

**【0102】**

(付記13) 前記他方のコントローラは、前記第2のキャッシュメモリのミラー領域へのリード／ライト処理を禁止するステップを更に有することを特徴とする付記12のストレージ制御方法。

**【0103】**

(付記14) 前記一方のコントローラが、前記依頼装置から複数ページ分のデータ書込み要求を受けたことに応じて、前記コピー中に、前記第1のキャッシュメモリのリード／ライト領域の割り付けられた次ページに次ページのデータを書込むステップを更に有することを特徴とする付記9のストレージ制御方法。

**【0104】**

(付記15) 前記一対のコントローラの各々は、前記キャッシュメモリ及び前記ストレージ装置を制御する制御ユニットと、前記コントローラ間での通信を行

うためのノードチャネル回路とを有することを特徴とする付記9のストレージ制御方法。

**【0105】**

(付記16) 前記コピーステップは、前記制御ユニットが、前記ノードチャネル回路に、前記第1のキャッシュメモリのリード／ライト領域の割り付けられたページのデータを、前記第2のキャッシュメモリのミラー領域の獲得ページにDMA転送することを指示し、前記コピーを行うステップからなることを特徴とする付記15のストレージ制御方法。

**【0106】**

**【発明の効果】**

このように、本発明では、相手コントローラのキャッシュメモリのミラー領域を割り当て、ミラー管理テーブルを作成することにより、コントローラ間でのミラーページ獲得メッセージのやり取りが不要となり、コントローラ間のハード、特に、CPUの負荷を軽減して、ライトデータのミラーリングが可能となる。しかも、相手コントローラのプログラム動作を不要とでき、コントローラ自体の性能を向上できる。

**【図面の簡単な説明】**

**【図1】**

本発明の一実施の形態のストレージシステムの構成図である。

**【図2】**

本発明の一実施の形態のミラー領域割り当て処理フロー図である。

**【図3】**

図2の本発明の一実施の形態のキャッシュメモリの説明図である。

**【図4】**

図3のミラー管理テーブルの構成図である。

**【図5】**

本発明の一実施の形態のデータ書き込み動作の説明図である。

**【図6】**

図5のデータ書き込み処理フロー図である。

**【図 7】**

図 1 のライトバック処理フロー図である。

**【図 8】**

図 7 のライトバック処理の動作説明図である。

**【図 9】**

図 1 のデグレード処理の説明図である。

**【図 1 0】**

本発明の他の実施の形態のデータ書込み処理フロー図である。

**【図 1 1】**

図 1 0 のデータ書込み処理の動作説明図である。

**【図 1 2】**

従来技術のデータ書込み処理の説明図である。

**【符号の説明】**

1、2 ストレージコントローラ

1 1、1 2、2 1、2 3 チャンネルアダプター

1 3、1 4、2 3、2 4 デバイスアダプター

1 0、2 0 C P U

1 5、2 5 メモリ

1 6、2 6 キャッシュメモリ

1 6 - 1、2 6 - 1 リード／ライト領域

1 6 - 2、2 6 - 2 ミラー領域

3 0、4 0 P C I - ノードブリッジ回路

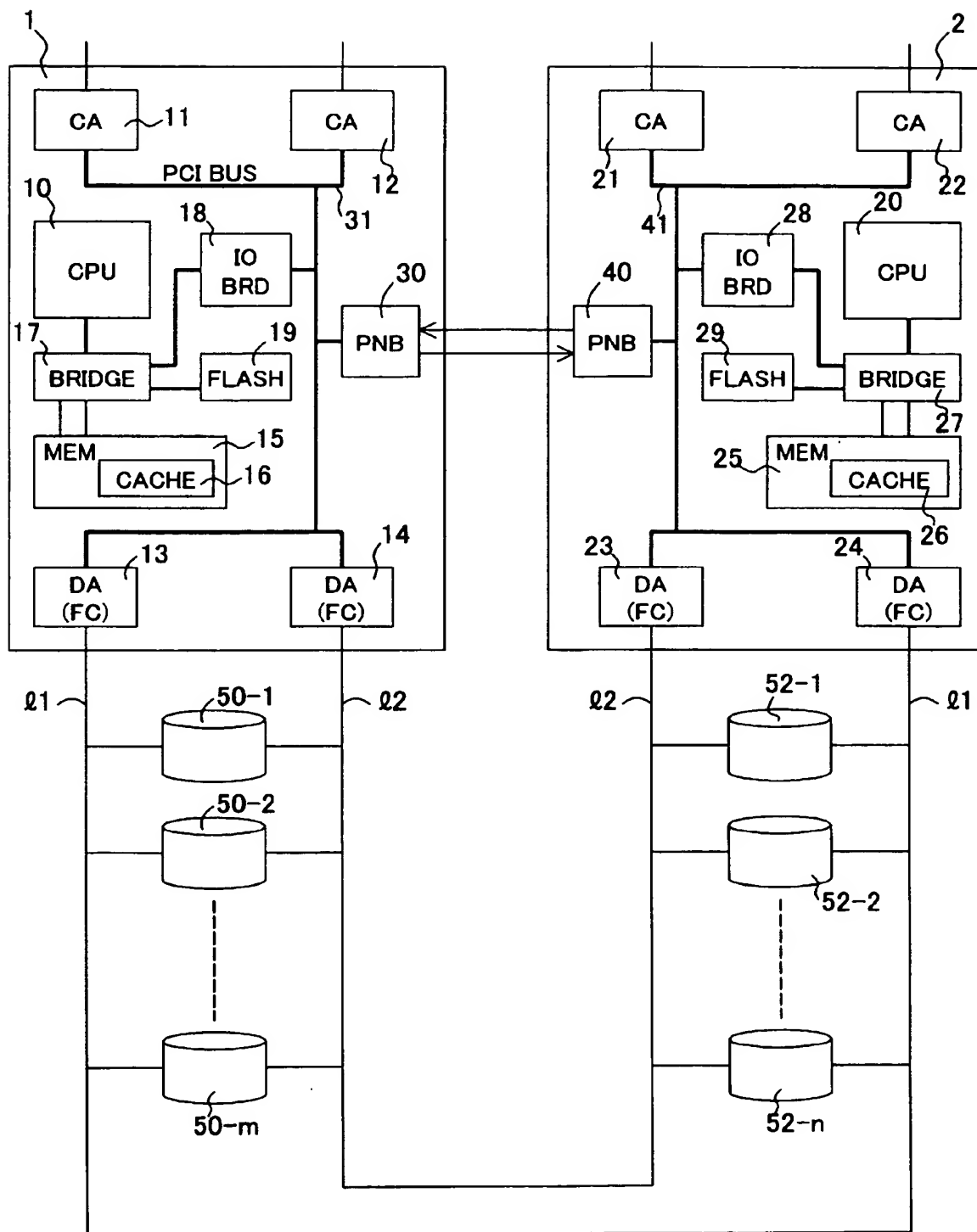
3 1、4 1 P C I バス

5 0 - 1 ~ 5 0 - m、5 2 - 1 ~ 5 2 - n 物理ディスク装置（ストレージ装置）

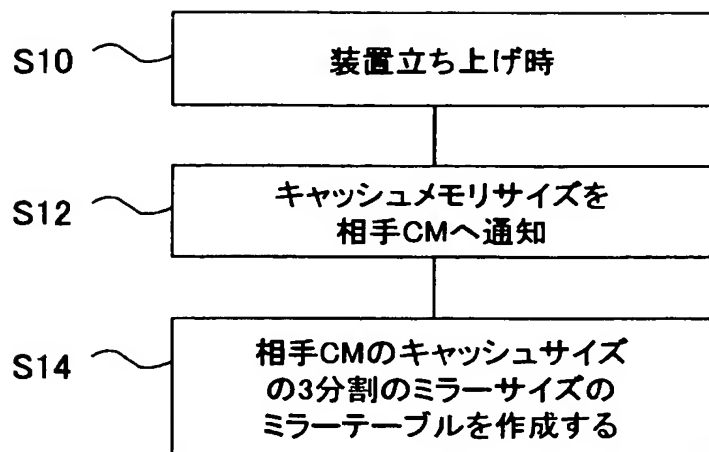
1 6 0、2 6 0 ミラー管理テーブル

【書類名】 図面

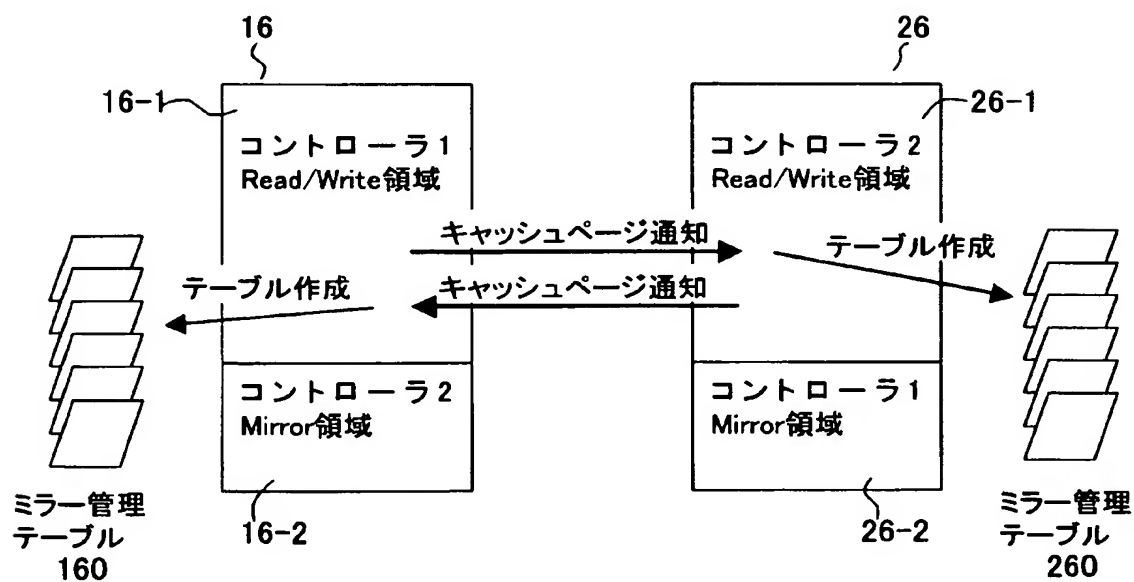
【図 1】



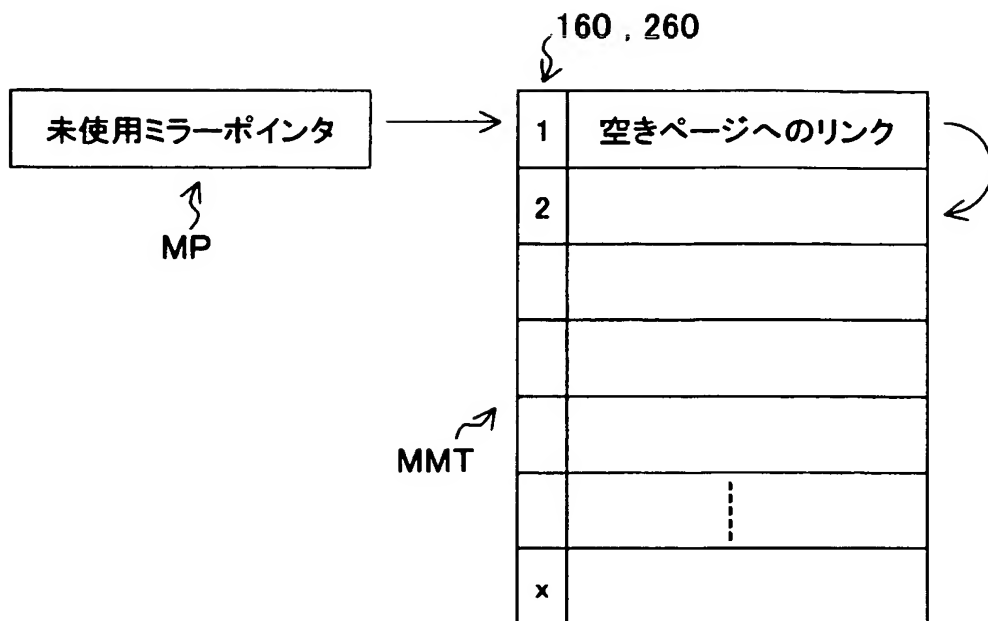
【図 2】



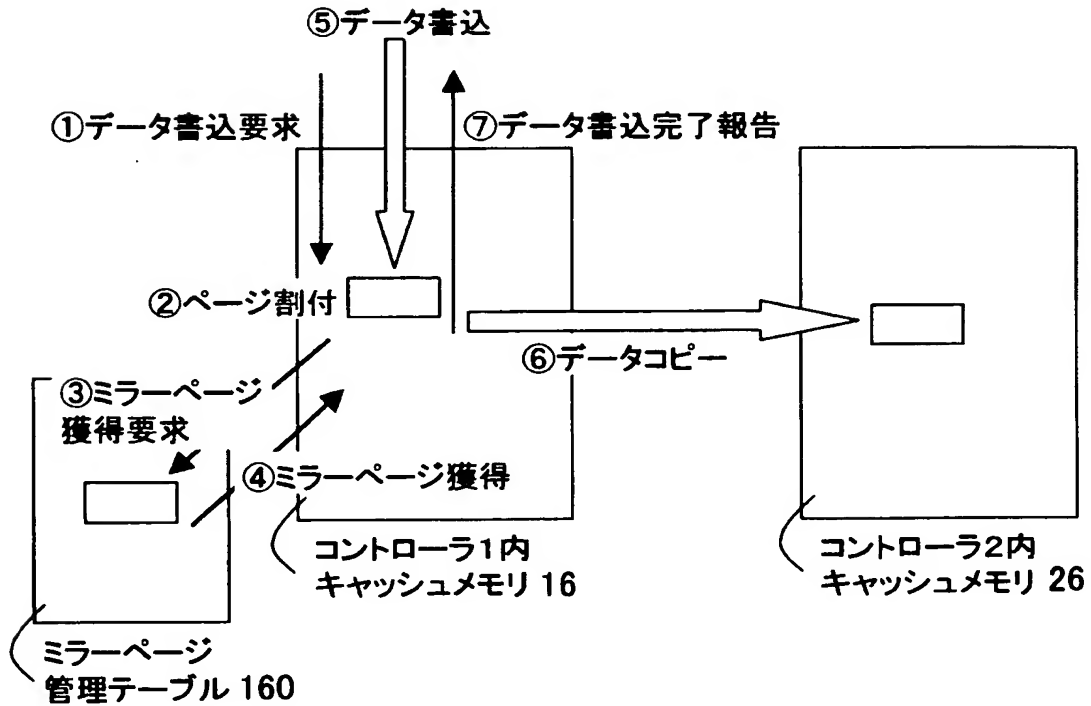
【図 3】



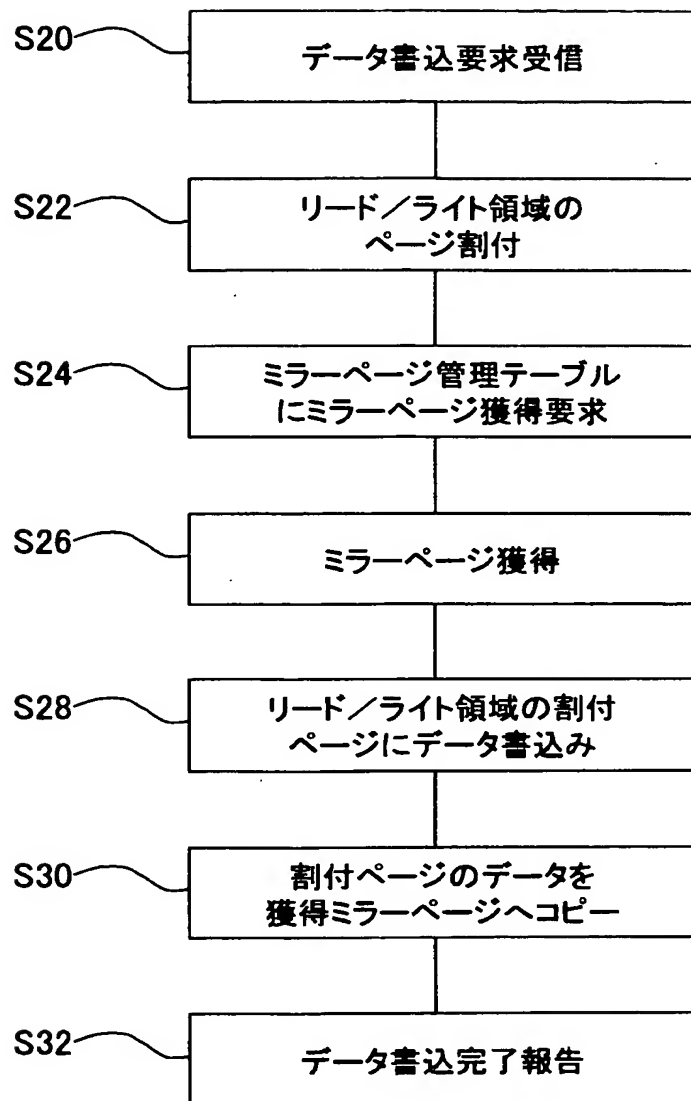
【図 4】



【図 5】

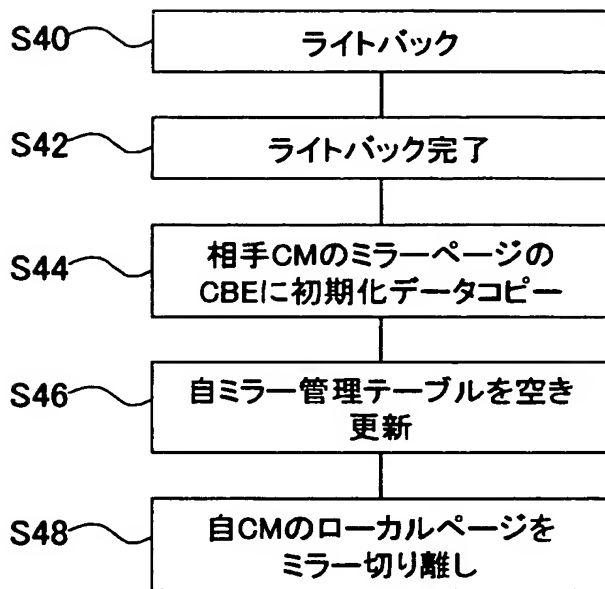


【図 6】

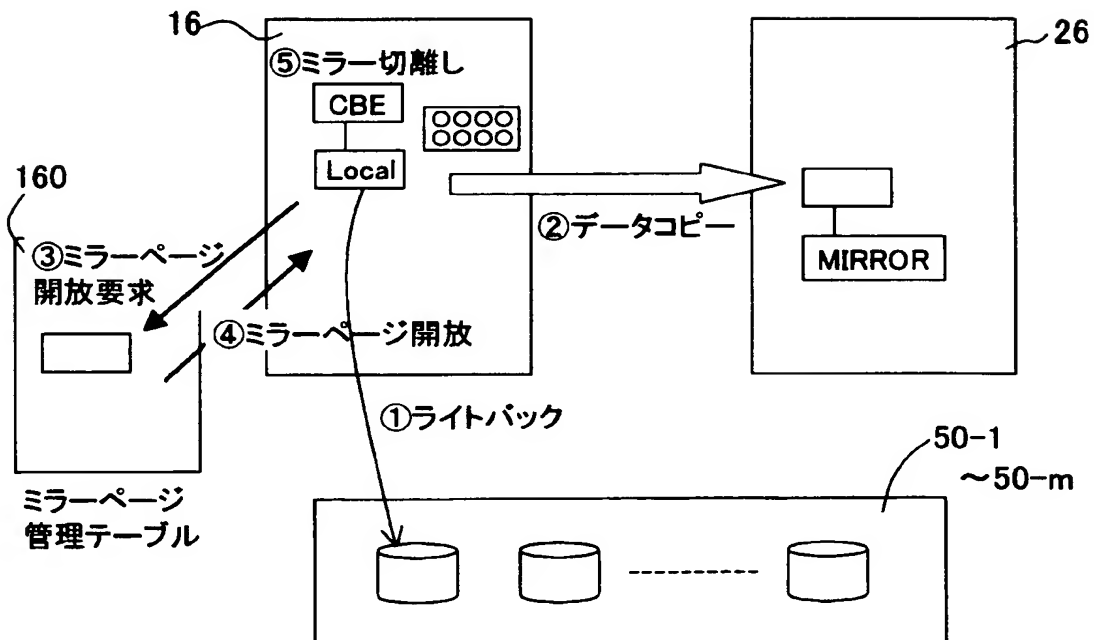




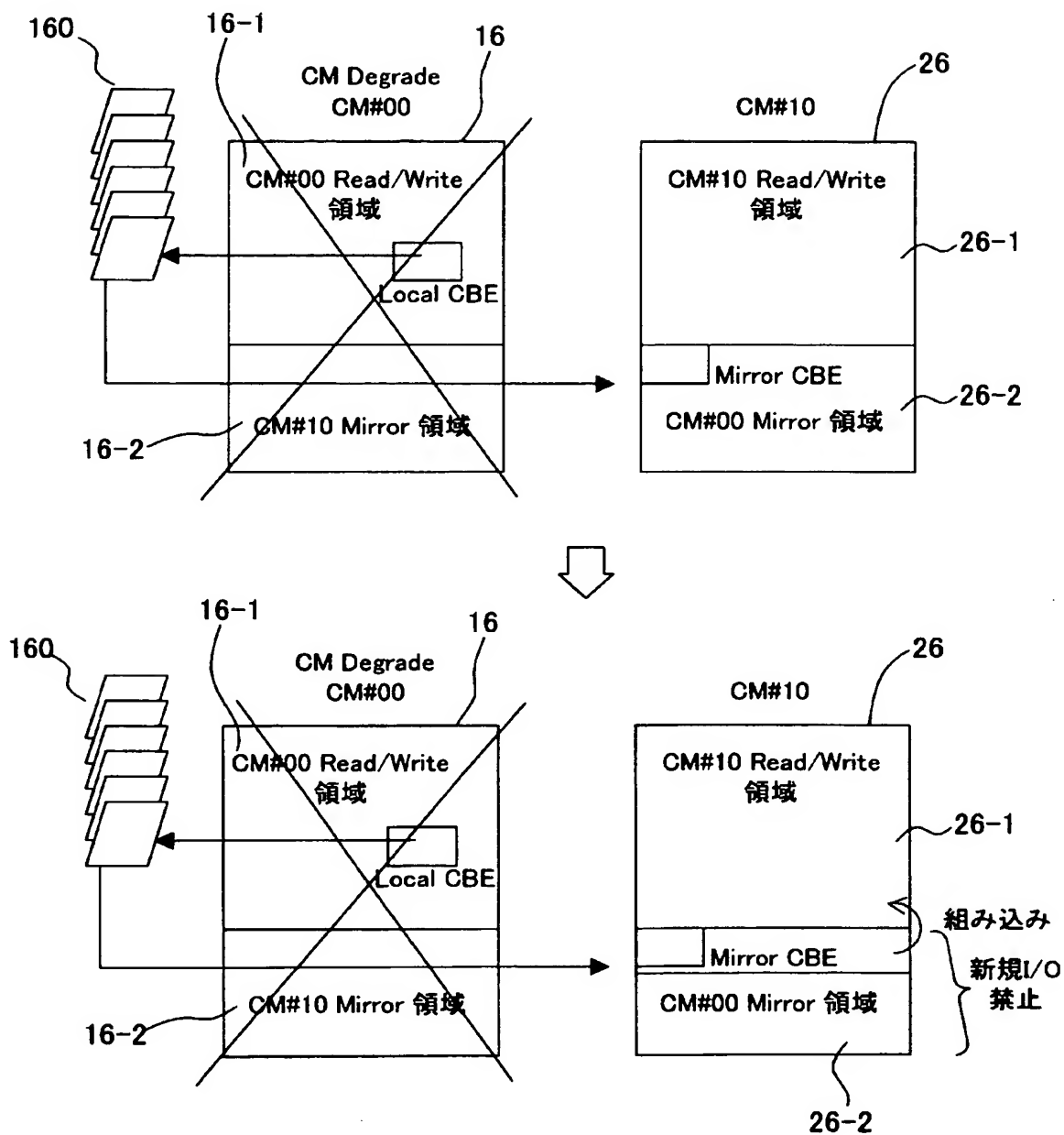
【図 7】



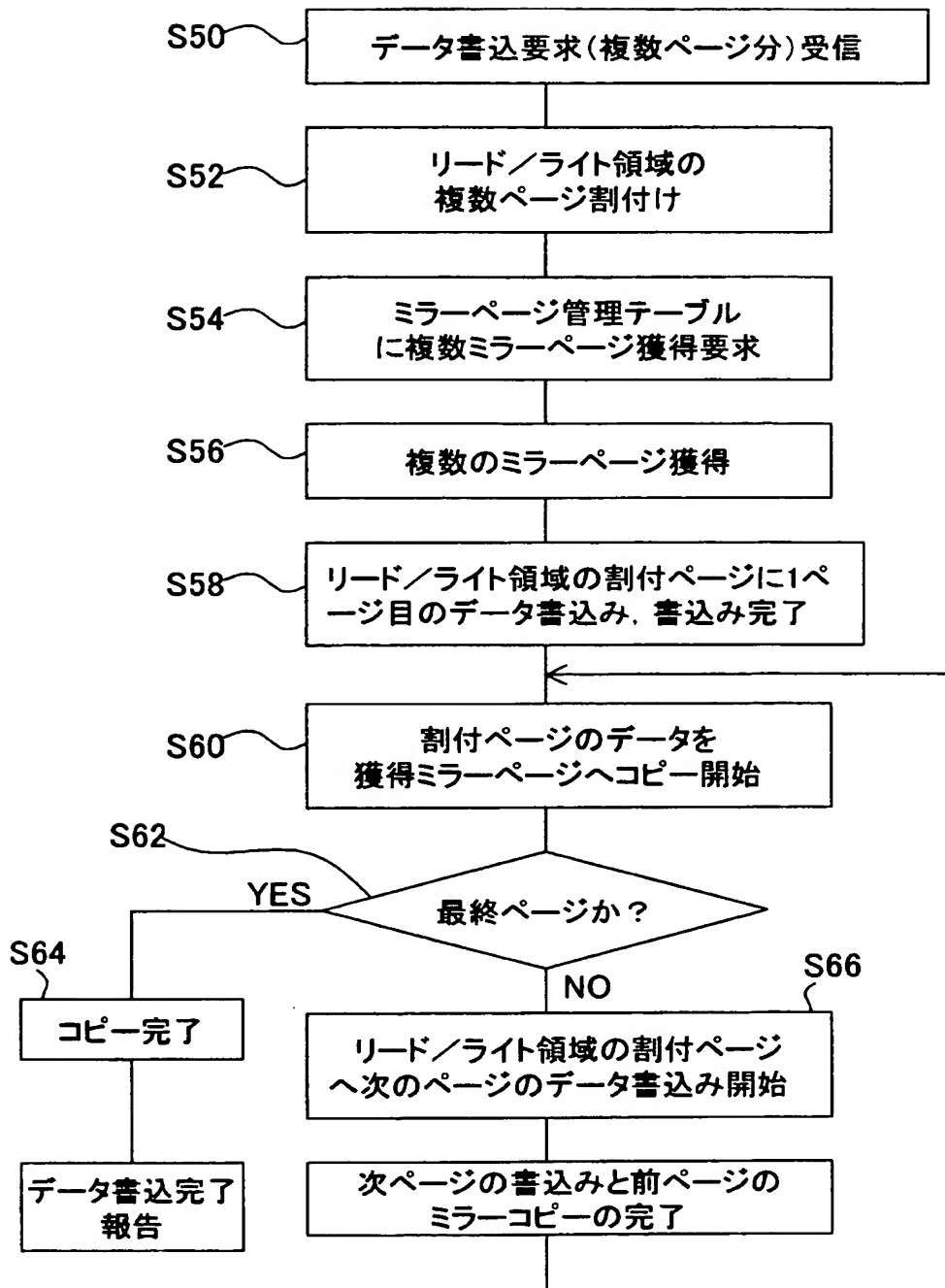
【図 8】



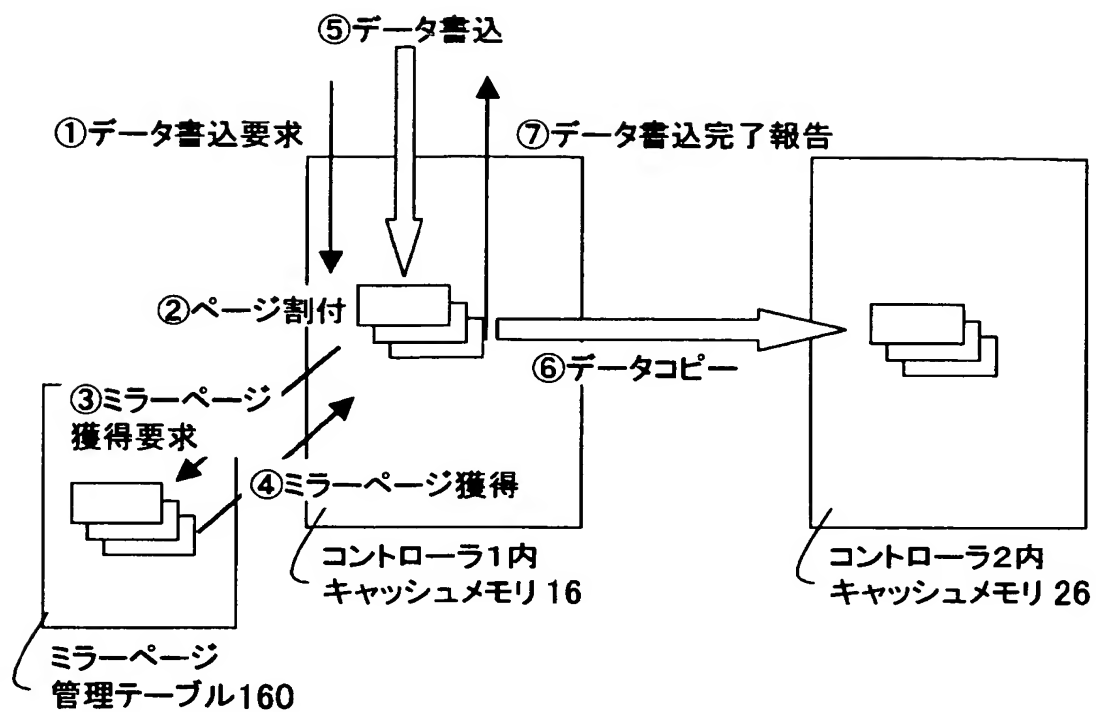
【図 9】



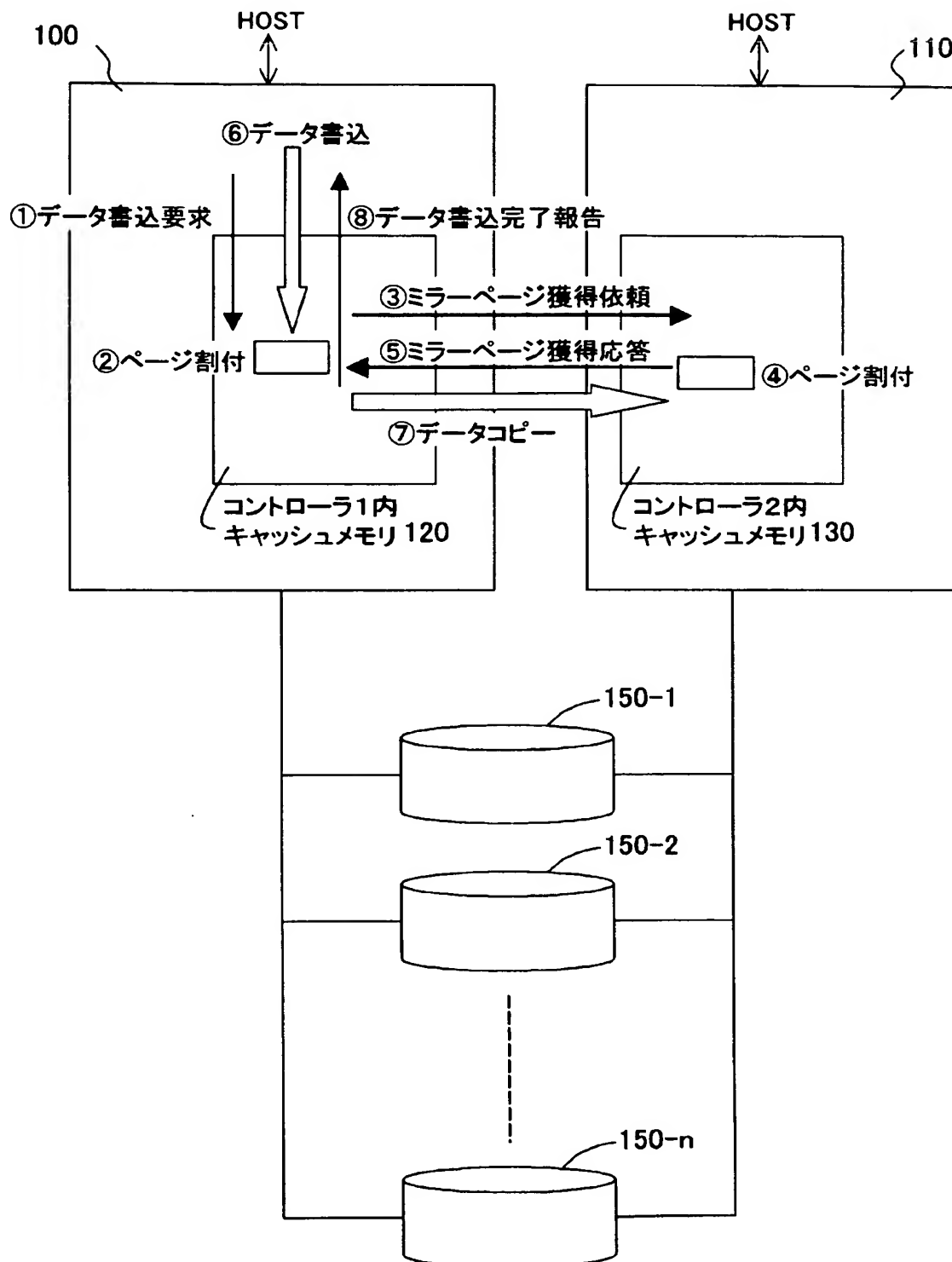
【図10】



【図 11】



【図 12】



**【書類名】 要約書****【要約】**

**【課題】** 一对のコントローラにより、ホストアクセスに応じて物理ディスク群を制御するストレージ制御装置において、キャッシュメモリへのデータライト時のミラーリング処理を軽減し、高速動作を可能とする。

**【解決手段】** 相手コントローラ（１、２）のキャッシュメモリ（１６、２６）のミラー領域（１６－２、２６－２）を割り当て、ミラー管理テーブル（１６０、２６０）を作成し、相手コントローラのキャッシュメモリのミラーページの獲得を、ミラー管理テーブルを参照して実行する。コントローラ間でのミラーページ獲得メッセージのやり取りが不要となり、コントローラ間のハード、特に、CPUの負荷を軽減して、ライトデータのミラーリングが可能となる。しかも、相手コントローラのプログラム動作を不要とでき、コントローラ自体の性能を向上できる。

**【選択図】** 図 5

特願 2 0 0 2 - 3 3 8 9 0 1

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 2 2 3 ]

1. 変更年月日

1 9 9 0 年 8 月 2 4 日

[変更理由]

新規登録

住 所

神奈川県川崎市中原区上小田中 1 0 1 5 番地

氏 名

富士通株式会社

2. 変更年月日

1 9 9 6 年 3 月 2 6 日

[変更理由]

住所変更

住 所

神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号

氏 名

富士通株式会社